Proceedings of ICMRED 2026 Symposium: The Future of Work: Strategy, Workforce Transformation, and Organizational Renewal
DOI: 10.54254/2754-1169/2026.LD32146

Data Management Strategies of Information Systems in the Big Data Era: Storage, Management and Value Maximization

Siqi Yin

School of Information Engineering, Shanxi Media College, Jinzhong, China
yinenyiii@163.com

Abstract. Big data technologies have caused the volume, velocity, and variety of data generated by enterprise information systems to grow exponentially. It has thereby rendered traditional approaches to data management no longer effective. This paper focuses on data storage, management, and analysis strategies for information systems within big data environments and searches for pathways to maximize value. Employing a combined literature review and case study methodology, it systematically reviews existing data management frameworks while integrating corporate practical experience through case studies. The findings indicate that data management activities of an existing information system experience problems, like storage resource management efficiency, the lack of standardized data governance, and inefficient use of analytical technologies. The proposed strategy involving data management hierarchies, data governance, and intelligent analytics can significantly increase the capacity of an information system in creating value from existing data. The research is beneficial and can offer significant help to existing and running enterprises on how they can improve data management practices and gain a competitive edge.

Keywords: Big data, Information systems, Data management strategy, Data storage, Data value

1. Introduction

The beginning of the Big Data era marks a paradigm shift in the world information system, fueled by the growth in data volumes, rates, and types—characterized by exponential growth in data volumes, ever-increasing data generation and processing rates, and burgeoning diversity in data types [1]. The challenge for information systems implemented in the past—to cope with immense volumes of neatly organized data—lies in adapting to the requirement for effective management of petabytes of data types such as structured, semi-structured, and unstructured data sourced from various places such as the Internet of Things (IoT) sensors, social media outlets, application systems, and multimedia data streams. However, such a paradigm shift has posed a huge gap between data management models and data systems [2]. While substantial research exists on discrete big data technologies, there remains a need for comprehensive frameworks that holistically address the synergistic integration of storage, governance, and analytics within cohesive enterprise IS architectures.

© 2026 The Authors. This is an open access article distributed under the terms of the Creative Commons Attribution License 4.0 (https://creativecommons.org/licenses/by/4.0/).

DOI: 10.54254/2754-1169/2026.LD32146

The specific research theme of this paper is the formulation and optimization of such integrated data management strategies for information systems in the big data environment. At the present stage of the research, the following are the core research questions that have been developed: what are the major architectural and managerial challenges currently affecting the field of IS DM as a result of the inherent properties that are inherent in big data; how can storage strategies be designed with regard to cost efficiency and performance; what are the elements of a successful approach to governing information at all stages of the information life cycle; and finally, how can the processes by which smart analytic approaches can be actually utilized and optimized with regard to the creation of maximum associated business value? In order to address these questions, the methodology blends a review of existing literature on big data theory and IS architecture, as well as data valuation, with a review focused on practical implementations through existing case study research.

2. Theoretical foundations of data management in information systems for the big data era

2.1. Characteristics and evolution of big data

Big data is conceptually defined by a set of attributes often encapsulated as the "5Vs": Volume, Velocity, Variety, Veracity, and Value.

"Volume" indicates the large amounts of information in which the projected global "datasphere" will ramp up to 181 zettabytes by 2025 [3]. Velocity indicates the speed at which information must be processed, often in near-real time. Variety indicates the kinds of information, which may be simple relational structures, JSON documents, images, video, or even location intelligence-like information types. Veracity relates to the uncertainties, noise, or biases that affect information veracity itself. Finally, Value indicates the information that has been processed from large amounts of raw information, creating useful insights, like intelligence in a traditional context, in which traditional approaches have been based upon hypothesis-driven, sample-based research, transforming into a comprehensive, exhaustive research pattern, often termed as a computer science revolution in traditional storage philosophies [4].

2.2. Architecture of information system and requirements for data management

Modern IS architecture has evolved from monolithic, application-centric architectures to distributed, modular, and often cloud-native microservices ecosystems. This evolution has brought about new and severe requirements for data management. Elastic Scalability: The ability to seamlessly accommodate unpredictable data growth; Polyglot Persistence: support for diverse data models with optimal storage solutions;

High Resilience and Availability, ensuring business continuity through replication and disaster recovery mechanisms; and Seamless Interoperability, facilitating data flow across hybrid and multi-cloud environments. Modern IS design increasingly decouples storage from compute resources, and separates operational systems (optimized for transactional latency) from analytics systems (optimized for complex query throughput), leading to layered architectures that include data lakes, data warehouses, and the lakehouse paradigm now emerging [5].

2.3. Data value theory and enterprise data strategy

Data is found to have no intrinsic value and the value rests upon the context and the use to which the data is put. The hierarchy proposed by the DIKW model describes the valuation or value creation as

DOI: 10.54254/2754-1169/2026.LD32146

a sequence that begins with the transformation of raw data into information by the addition of context, into knowledge by the perception of patterns and relationships, and finally into the level of wisdom that facilitates action [6].

The Infonomic approach, as formally articulated by Douglas Laney, explicitly supports this approach by encouraging the consideration of information as an economic asset in which strategic investments must be made [7]. It follows, therefore, that a comprehensive approach to the data strategy must, explicitly and directly, be linked with a corporate goal, a desire to improve the customer experience, a need to optimize the entire chain, navigate financial risk, etc., to elevate data management to a critical value-generating activity.

3. The present situation and challenge of information system data management under big data's environment

3.1. Analysis of the current state of enterprise data management

Many organizations face a reality of "data sprawl", where information resides in fractured states in legacy on-premise solutions, various organizational databases, differing Software as a Service (SaaS) entities, and assorted cloud storage repositories. These fractured states lead to a reality of deep-seated silos of information that serve to fragment a singular state of trust in organizational information. According to a recent industry survey, a significant majority of organizations report being impeded by data silos, while less than half claim to be effectively competing based on data and analytics [8]. Data quality in such silos often follows inconsistent patterns, and generally, metadata management is either absent or inadequate. This makes it very difficult for anybody to effectively find, have confidence in, and rightly interpret available data assets.

3.2. Impact of big data on information systems

The big data phenomenon has necessitated a fundamental re-architecting of information systems. Traditional Relational Database Management Systems (RDBMS), while excellent for ACID-compliant transactional workloads, often struggle with the scale, flexibility, and cost requirements for semi-structured data and real-time ingestion streams. This technological pressure has led to the proliferation of NoSQL databases and dedicated stream-processing platforms. Moreover, the pure financial expense of storing all created data on high-performance media became too prohibitive to further scale, and a strategic shift towards intelligent, policy-driven data tiering was required. The analytics workload has also moved from scheduled, backward-looking batch reporting to interactive exploration, machine learning model training, and real-time inference.

3.3. Key challenges

3.3.1. Low efficiency and high consumption of storage system

One-size-fits-all storage mode is widely adopted in the industry, which keeps cold data with low access frequency in high-cost primary storage, resulting in a large amount of unnecessary capital loss; coupled with the problem of redundancy and duplication caused by data isolated islands, it further pushes up storage costs and increases management complexity.

DOI: 10.54254/2754-1169/2026.LD32146

3.3.2. Data governance and quality absence

At the enterprise level, there is a lack of a standardized global data governance framework, which leads to a vague definition of data ownership, an inconsistent business definition and no monitoring of data quality, which not only weakens data credibility, but also hinders compliance and greatly increases operational risk.

3.3.3. Low maturity of analytical ability and unclear return on investment

There is a significant gap between data collection and value landing, and the core pain points include the global shortage of professional data talents, the high complexity of analysis tools, making it difficult for business personnel to use, and the disconnection between analysis projects and the core business needs of the enterprise. Finally, it is difficult to quantify the return on investment.

3.4. Case study analysis

An examination of technology giants like Alibaba Group provides a clear example of a coordinated response to these challenges. To manage colossal and diverse data, Alibaba pioneered the "Data Middle Office" concept.

This initiative involved architecturally and organisationally dismantling the data silos between various data domains, which were diverse in terms of the business units involved, in order to establish a unified, standardised, single data solution as a layer throughout the entire enterprise. This was the starting point, which also enabled the direct enhancement of the organisation concerning the personalisation of the customers.

4. Information system storage strategies

4.1. Multi-tiered storage architecture with different types of data

It also requires the development of a forward-thinking, financially sound storage strategy, which includes the need for a tiered architecture, creating a strong congruence between the cost of storage, access, performance, and value.

Hot/Performance Tier: Designated for mission-critical latency-prone data with microsecond access requirements. It typically makes use of premium Solid-State Drives (SSDs) or cloud block storage.

Cool/Standard Tier: Accessing stored information periodically according to reporting needs, analysis, or development practices, this tier utilizes cost-effective storage devices, which provide an optimal trade-off between cost and access latency [9].

Cold/Archive Tier: Reserved for data whose primary retention is used for compliance and is normally not accessed. This tier uses very low-cost tape library systems or cloud archive services, where retrieval times can range from minutes to several hours.

Technologies like Hadoop HDFS for large-scale analytics and cloud object storage with automated lifecycle policies are inherently engineered to support and automate this tiered model efficiently [10].

DOI: 10.54254/2754-1169/2026.LD32146

4.2. Elastic storage technologies

Cloud computing has made elasticity a cornerstone of the modern storage strategy. Services like Amazon S3, Azure Blob Storage, and Google Cloud Storage provide seemingly endless on-demand scale, eliminating the need for expensive and speculative over-provisioning of physical hardware. Moreover, it enables serverless data platforms to decouple storage and compute resources such that each can scale independently and precisely based on the exact workload. Such is an architectural pattern that optimizes both performance and cost.

4.3. Balancing storage costs and performance

To achieve an optimal balance, active data management policies and techniques need to be implemented: (1) Automated Data Lifecycle Management: Developing and applying policies that automatically move data between storage tiers according to specified criteria. (2) Data Optimization Techniques: Using techniques such as compression and deduplication to minimize storage capacity requirements as much as possible. (3) Query Performance Engineering: Leverage the principle of efficient data partitioning, data indexing, and caching to optimize query performance for analytical operations. To gain an accurate knowledge of the Total Cost of Ownership for any storage environment, the evaluation must go beyond the basic costs of the physical media involved to include an evaluation.

5. Development and optimisation of information system data management frameworks

5.1. Establishing data management frameworks

A framework is vital if the organization is to take relevant and effective action. The DAMA-DMBOK2 model is recognized as the framework to be followed when building the structure of the data management function. DALMA-DMBOK2 is particularly advantageous as the framework is composed of the following knowledge areas: data governance, data architecture, data quality, and data security, thus creating the full picture of how to effectively create the data management function [11]. Creation of the cross-functional data governance council, comprising leaders from the business side as well as the IT side of the business. They will define the high-level data policies as well as create the centralized business glossary to ensure the entire organization understands the definitions of the key entities within the business.

5.2. Full lifecycle management

For effective data governance to take place, it has to cover all the stages in the lifecycle of the data, namely:

Plan & Create: Standards for what constitutes data capture, metadata assignment, and classification at the point of origin.

Store & Process: Data quality rules, data transformations, or tiering rules are applied here.

& Analyze: Role-based access controls, enforcing usage patterns, use of detailed information as it relates to its origins, or its lineage, etc.

Archive & Destroy: Data retention schedules must be implemented and executed in such a manner that is legally defensible for deletion once the data's lifecycle ends.

DOI: 10.54254/2754-1169/2026.LD32146

5.3. Ensuring data security and compliance

Data security is a non-negotiable pillar. A defense-in-depth approach is required, incorporating encryption for data both at rest and in transit, implementing fine-grained access controls, and maintaining comprehensive audit logging. A "privacy by design" philosophy must be integrated, employing techniques like data masking and anonymization to comply with stringent global privacy regulations such as the GDPR. Frameworks like the NIST Cybersecurity Framework offer structured, widely adopted guidance for protecting information assets [12].

6. Approaches to maximizing data value---construction of data analysis platform

The platform for value extraction is a modern data analytics architecture, often conceptualized in layers: a Data Ingestion Layer for collecting data from various sources; a Storage & Processing Layer comprising a data lake and/or a data warehouse powered by distributed processing engines like Apache Spark; a Semantic & Modeling Layer where business logic and metrics are defined; and an Analytics & Visualization Layer serving data scientists and business users via Bl platforms. The trend is toward fully managed cloud platforms that simplify infrastructure management and allow organizations to focus resources on analytics rather than administration.

7. Conclusion

This research systematically explores the necessity of advanced, integrated data management strategies in enterprise information systems to unlock big data's transformative potential. Analysis reveals that these foundational characteristics of big data management—volume, velocity, and variety—make traditional segregated and monolithic models of management less relevant in this context [1,3]. The interrelated systemic challenges identified include inefficient repository models combined with high storage costs, weak organizational governance models, and analytical immaturity.

Findings affirmatively validated the indisputable need to have an essential holistic strategic model formed of the three pillars of the model. First, the need to have an adaptive multi-layered storage infrastructure to accommodate the rational mapping of the Enterprise Data Assets to the best media type of storage according to access patterns and economic value.Second, the need to have a full lifecycle data governance framework. Third, the strategic need to have an "intelligent" scalable paradigm to leverage the "power of machine learning to insights".Assessing dominant business practices in a leading corporation also empirically justifies our aforesaid model by considering its efficiency in achieving hyper-personalization, risk mitigation, and overall business excellence.

Finally, maximizing data value is a strategic journey that continues into perpetuity, necessitating sustained investment in technology, a data-literate culture, and agile organizational structures that can take quick action on generated insights.

References

[1] Gandomi A., Haider M. (2015) Beyond the hype: Big data concepts, methods, and analytics. International Journal of Information Management, 35(2): 137-144.
[2] Chen M., Mao S., Liu Y. (2014) Big Data: A Survey. Mobile Networks and Applications. 19(2): 171-209
[3] Reinsel D., Gantz J., Rydning J. (2021) The Digitization of the World: From Edge to Core. IDC White Paper.
[4] De Mauro A., Greco M., Grimaldi M. (2016) A formal definition of Big Data based on its essential features. Library Review. 65(3): 122-135.

DOI: 10.54254/2754-1169/2026.LD32146

[5] Armbrust M., et al. (2021) Lakehouse: A New Generation of Open Platforms that Unify Data Warehousing and Advanced Analytics. CIDR2021.

[6] Rowley J. (2007) The wisdom hierarchy: representations of the DIKW hierarchy. Journal of Information Science. 33(2): 163-180.

[7] Laney D.B. (2017) Infonomics: How to Monetize, Manage, and Measure Information as an Asset for Competitive Advantage. New York: Bibliomotion.

[8] New Vantage Partners. (2023) Big Data and Al Executive Survey 2023. so.com/link?m=bnmTBhNAhgja582alb0fvTOYsqkrJkr2ad5MnH%2BOQEkQ7rFmtR%2FaDbHCWf%2BykTrxJOe0C%2FQqn70TEk9JJ%2FuiLdgDWaW6ip7TARyjdZfB7q6zwGnuxZlRZOCRHlN63LPRUbgGYJhDSzb32%2BoxvlOzQ4ctd0elZ0Y4qOk9xYmA50jc8wZxgLHiUZifa22Pfuj8H%2B3WXiQ%3D%3D

[9] Alibaba Cloud. (2022) The Alibaba Data Middle Office: A Strategic Approach to Unify Data Assets. https://www.aliyun.com

[10] A.L. Jawetz. (2024) Arquitetura serverless para sistema de gerenciamento de atendimentos psicológicos. UNIVERSIDADE FEDERALDORIOGRANDEDOSUL. lume.ufrgs.br

[11] DAMA International. (2017) DAMA-DMBOK: Data Management Body of Knowledge. 2nd ed. New Jersey: Technics Publications.

[12] National Institute of Standards and Technology (NIST). Framework for lmproving Critical Infrastructure Cybersecurity, Version 1.1.2018.